

# Ergodicity of Eigenfunctions for Ergodic Billiards

Steven Zelditch, Maciej Zworski

Mathematics Department, Johns Hopkins University, Baltimore, MD21212-2689, USA

Received: 15 November 1994 / in revised form: 10 February 1995

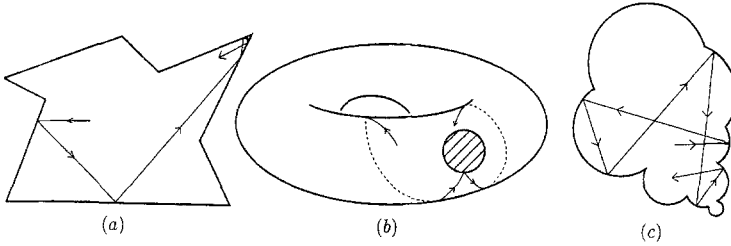
**Abstract:** We give a simple proof of ergodicity of eigenfunctions of the Laplacian with Dirichlet boundary conditions on compact Riemannian manifolds with piecewise smooth boundaries and ergodic billiards. Examples include the “Bunimovich stadium”, the “Sinai billiard” and the generic polygonal billiard tables of Kerckhoff, Masur and Smillie.

## 1. Introduction and Statement of Results

The purpose of this note is to give a simple proof of ergodicity of eigenfunctions on manifolds with piecewise smooth boundaries and ergodic billiard flows. For the boundaryless case this is now well known and was established in [14, 17, 3]. If the manifold is a convex region in  $\mathbb{R}^n$  with  $W^{2,\infty}$  boundary, the same result was proved much more recently by Gérard and Leichtnam [6] who needed, among other things, a quantitative version of the results of Melrose and Sjöstrand on propagation of singularities for mixed problems (see [12 and 9, Chap. 24]). Our argument uses an aspect of an abstract approach to quantum ergodicity developed by the first author in [18] and some basic microlocal analysis on manifolds with boundary. It applies to a class of (not necessarily convex) regions with measure theoretically negligible sets of singular points. That class includes the much studied example of the “Bunimovich stadium” and other examples of non-dispersive billiard tables [1], the “Sinai billiard” [15] as well as the ergodic polygonal billiard tables of Kerckhoff, Masur and Smillie [11] – see Fig. 1.

Referring to Sect. 2 for precise definitions of intuitively clear notions of manifolds with piecewise smooth boundaries and of billiard flows, we can state our result as

**Theorem.** *Let  $(M, g)$  be a compact  $\mathcal{C}^\infty$  Riemann manifold with piecewise smooth boundary. Let  $\{\phi_j\}_{j \in \mathbb{N}_0}$  be an orthonormal set of eigenfunctions of the Laplacian,  $\Delta_g$ , on  $M$ . If the billiard flow defined on a set of full measure in  $S^*M$  is ergodic then there exists a subsequence  $\{j_k\}$  of density 1,  $\#\{j_k : j_k \leq N\}/N \rightarrow 1$*



**Fig. 1.** Examples of nonconvex ergodic billiards: (a) *generic polygon*, (b) *Sinai billiard*, (c) *non-dispersive billiard table*.

such that

$$\langle A\phi_{jk}, \phi_{jk} \rangle \xrightarrow[k \rightarrow \infty]{} \int_{S^*M} \sigma_0(A) d\mu \tag{1.1}$$

for any  $A \in \Psi_{\text{phg}}^0(M)$  with Schwartz kernel compactly supported in  $M^\circ \times M^\circ$ . Here,  $\sigma_0(A)$  denotes the principal symbol of  $A$ .

By applying standard heat equation methods, one could extend this to a less restrictive class of operators as in [6], removing the compact support condition on the kernel. To keep the presentation simple we will carry this out for multiplication operators only:

**Corollary.** *With the notation and assumptions of the theorem above*

$$|\phi_{jk}|^2 \xrightarrow[k \rightarrow \infty]{} 1 \text{ weakly in } L^1(M, d\text{vol}_g). \tag{1.2}$$

### 2. Preliminaries

The billiard tables we consider are compact  $\mathcal{C}^\infty$  Riemann manifolds  $(M, g)$  with piecewise smooth boundaries in the sense of Cornfeld–Fomin–Sinai, [4], Chapter 6. Thus we will assume that  $M$  is a compact subset of a  $\mathcal{C}^\infty$  compact manifold  $\tilde{M}$  of the same dimension and that there exist  $r$  functions  $f_1, \dots, f_r \in \mathcal{C}^\infty(\tilde{M})$ , such that

- (i)  $df_i \upharpoonright_{f_i^{-1}(0)} \neq 0$ ,
  - (ii)  $df_i$  and  $df_j$  are independent on  $f_i^{-1}(0) \cap f_j^{-1}(0)$ ,
  - (iii)  $M = \{x \in \tilde{M} : f_i(x) \geq 0, 1 \leq i \leq r\}$ .
- (2.1)

Condition (ii) which is taken directly from [4] can be replaced by a weaker condition allowing an inclusion of Bunimovich’s stadium in our class of billiard tables:

- (ii)'  $f_i^{-1}(0) \cap f_j^{-1}(0)$  is an embedded submanifold of  $\tilde{M}$

and  $M$  has a Lipschitz boundary,

where a boundary is Lipschitz if it is given locally by a graph of a Lipschitz function (we need that assumption to easily quote the facts about the Dirichlet problem). We assume that the metric  $g$  extends to a smooth Riemann metric on  $\tilde{M}, \tilde{g}$ .

The boundary  $\partial M$  is the union  $\bigcup_{i=1}^r \partial M_i$  with  $\partial M_i = f_i^{-1}(0) \cap M$ . The *regular* components are given by  $\partial \tilde{M}_i = \partial M_i \setminus \bigcup_{k \neq i} \partial M_k$  and we obtain the regular part of the boundary:  $\partial M_{\text{reg}}$ ; its elements are called *regular* points. The set  $S = \partial M \setminus \partial M_{\text{reg}}$  is said to consist of *singular* points.

The billiard flow  $\Phi^t$  is a (discontinuous) flow on a subset  $S^*M \setminus E$  of full measure in the unit cotangent bundle or tangent bundle – in what follows we will freely identify vectors and co-vectors using the metric. The flow  $\Phi^t$  is very closely connected to the generalized bicharacteristic flow of  $p = \tau^2 - g$  on  $\{p = 0\} \subset T^*(\mathbb{R} \times M) \setminus 0$  defined for  $\partial M$  smooth in [12] – see also [9], Sect. 24.3. However, the two are not the same as regards the billiard trajectories which touch  $\partial M_{\text{reg}}$  tangentially. Since the generalized bicharacteristic flow, rather than  $\Phi^t$ , arises naturally in the study of eigenfunctions we will have to introduce a common subsystem of the two.

Let us first recall, following [4] closely, the definitions of the billiard flow  $\Phi^t$  and of its phase space. For  $(x, v) \in S^*M^\circ$  the trajectory of  $\Phi^t(x, v)$  is the geodesic of  $(\tilde{M}, \tilde{g})$  through  $(x, v)$  until it intersects the boundary in the base, at a point  $y$ , say. If  $y \in M_{\text{reg}}$ , then the tangent vector on the geodesic,  $(y, w)$  will lie in the set  $\partial S^*M^- = \{(x, v) : x \in \partial M_{\text{reg}}, g_x(v, n_x) \leq 0\}$  of outgoing vectors, where  $n_x$  is the unit normal vector to  $\partial M$  at  $x$ . It is then reflected into the set  $\partial S^*M^+ = \{(x, v) : x \in \partial M_{\text{reg}}, g_x(v, n_x) \geq 0\}$  by the reflection  $\mathcal{R}(y, w) = (y, w - 2g_y(w, n_y)n_y)$ , and the flow then follows the geodesic through  $\mathcal{R}(y, w)$ .

From the dynamical point of view there are three *bad* sets in  $S^*M^\circ$ : (i) the set  $N_1$  of  $(x, v) \in S^*M^\circ$  whose trajectories hit  $S$  at some time; (ii) the set  $N_2$  of  $(x, v) \in S^*M^\circ$  whose trajectories reflect infinitely often in finite time; (iii) the set  $I$  of  $(x, v) \in S^*M^\circ$  whose forward or backward trajectories do not intersect the boundary. From the microlocal point of view there is an additional *bad* set for which the standard Egorov's theorem fails to hold: (iv) the set  $\mathcal{F}$  of  $(x, v) \in S^*M^\circ$  whose trajectories touch  $\partial M_{\text{reg}}$  tangentially at some time.

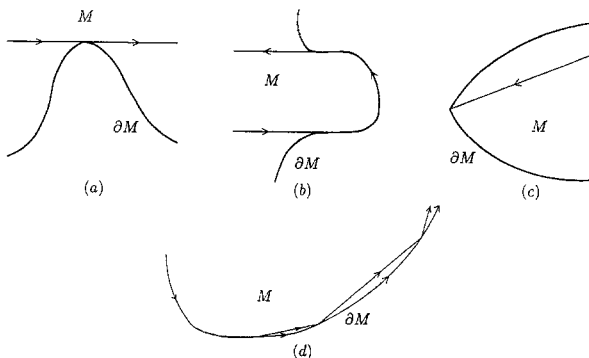


Fig. 2. Deleted rays: (a) grazing, (b) gliding along the boundary, (c) singular, (d) non-unique.

It is proved in [4], Sect. 6.2, that  $\mu(N_1) = \mu(N_2) = 0$ , where  $\mu$  is the Liouville measure on  $S^*M$ . (Strictly speaking we assume (i)–(iii) of (2.1); the argument of the proof applies however under the weaker assumption (ii)' or in fact just the first part of it.) It is also assumed that the billiards are *proper* in the sense that  $\mu(I) = 0$ . This is a natural assumption in the study of ergodic billiards since  $I$  is an invariant set. The phase space of  $\Phi^t$  is then defined as  $S^*M \setminus (N_1 \cup N_2 \cup I)$  in [4]. The phase space we will adopt here is

$$\mathcal{P} \stackrel{\text{def}}{=} S^*M \setminus (N_1 \cup N_2 \cup I \cup \mathcal{F}). \tag{2.2}$$

Since  $\mathcal{F}$  is also invariant it is natural to assume that  $\mu(\mathcal{F}) = 0$  in the study of ergodic billiards. However, this is no additional assumption since we have

**Lemma 1.** *The set*

$$\mathcal{F} = \{m \in S^*M^\circ : \exists t \in \mathbb{R} | t^* \Phi^t(m) \in S^*(\tilde{\partial M}_{\text{reg}})\}, \quad \iota : \partial M \hookrightarrow M, \tag{2.3}$$

satisfies  $\mu(\mathcal{F}) = 0$ .

*Proof.* We follow closely the argument of [4], Sect. 6.2 and refer there for unexplained notation and background. Thus, let  $d\mu_1$  be the surface measure on  $S^*_{\partial M_{\text{reg}}} M$  defined on  $S^*_{\partial M_i} M$  by

$$d\mu_1(x, v) = d\rho_i d\omega_x(v) |g_x(n_x, v)|,$$

where  $d\rho_i$  is the volume density of the metric  $g_i$  induced by  $g$  on  $\partial \tilde{M}_i$  and where  $d\omega_x$  is the usual Euclidean surface measure on  $S^*_x(\partial \tilde{M}_i)$ . Also, let  $\mathcal{M} \subset \mathbb{R} \times \partial S^*M^+ = \{(t, x, v) : 0 < t \leq f(x, v)\}$ , where  $f(x, v)$  is the length of the geodesic segment determined by the incoming vector  $(x, v)$ , that is the time at which the geodesic first intersects the boundary again. Since  $\mu(I) = 0$ ,  $f(x, v) < \infty$  except on a set of measure zero. Let us define the map  $\Phi : (t, x, v) \mapsto \Phi^t(x, v)$ ,  $\Phi : \mathcal{M} \rightarrow S^*M$ . The map is clearly onto  $S^*M \setminus I$ . Moreover, we have an equality of volume forms:  $\Phi^*(d\mu) = dt \otimes d\mu_1$  – see [4]. Hence if  $E \subset S^*M \setminus I$ ,

$$\mu(E) \leq \int_{\partial S^*M^+} \left( \int_0^{f(x,v)} \Phi^* \chi_E \right) d\mu_1. \tag{2.4}$$

To show that  $\mu(\mathcal{F}) = 0$  we first decompose  $\mathcal{F}$  as follows:  $\mathcal{F} = \bigcup_{n \in \mathbb{N}_0} \mathcal{F}_n$ , where  $\mathcal{F}_n$  is the set of  $(x, v)$  whose billiard flow trajectory first intersects  $\partial M_{\text{reg}}$  tangentially at the  $n^{\text{th}}$  time of intersection with it. We also define

$$\begin{aligned} \mathcal{F}_{nm}^\partial &= \{(x_-, v_-) \in \partial S^*M^+ : f(x_-, v_-) \leq m \\ &\text{and } \exists 0 < t < f(x_-, v_-), \Phi^t(x_-, v_-) \in \mathcal{F}_n\}, \end{aligned}$$

and  $\mathcal{F}_n^\partial = \bigcup_{m \in \mathbb{N}} \mathcal{F}_{nm}^\partial$ , so that we have  $\mathcal{F}_{n+1}^\partial = B^n(\mathcal{F}_1^\partial)$ , where  $B$  is the billiard ball map defined almost everywhere on  $\partial S^*M^+ - [4, 8]$ . Using (2.4) we have

$$\mu(\mathcal{F}) \leq \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} m \mu_1(\mathcal{F}_{nm}^\partial).$$

Since  $\mu_1$  degenerates on  $\mathcal{F}_1^\partial$  and since  $B_* \mu_1 = \mu_1$  (see [4, 8]) we have  $\mu_1(\mathcal{F}_{nm}^\partial) \leq \mu_1(\mathcal{F}_n^\partial) = \mu_1(\mathcal{F}_1^\partial) = 0$ . Hence  $\mu(\mathcal{F}) = 0$ .  $\square$

The lemma shows that the phase space (2.2) has full measure. To apply analytic techniques we need a larger set which will depend on a parameter  $T$ . Thus we define

$$X_T = \{(x, v) \in S^*M_{\text{reg}} : \exists \varepsilon(x, v) > 0 \forall |t| \leq T d_g(S, \pi \circ \Phi^t(x, v)) > \varepsilon(x, v) \\ \text{and } \iota^*(\Phi^t(x, v) \cap S^*_{\partial M_{\text{reg}}} M) \cap S^* \partial M_{\text{reg}} = \emptyset\}, \tag{2.5}$$

where  $d_g$  is the Riemannian distance with respect to  $g, \pi : S^*M \rightarrow M$  is the natural projection and  $\iota : \partial M \mapsto M$ . It is clear that there is no problem in defining  $\Phi^t, |t| \leq T$  on  $X_T$ .

**Lemma 2.** *The set  $X_T$  is open and has full measure in  $S^*M$ .*

*Proof.* Let  $(x, v) \in X_T$ . We start by observing that there exists at most finitely many  $t$ 's :  $t_1, \dots, t_k$  such that  $\Phi^t(x, v) \in S^*_{\partial M_{\text{reg}}} M$  and  $|t| \leq T$ . Otherwise for the limit point  $\bar{t}$  of infinitely many such  $t$ 's we would either have  $\lim_{t_j \rightarrow \bar{t}} \pi \circ \Phi^{t_j}(x, v) \in S$  or  $\iota^* \Phi^{\bar{t}}(x, v) \in S^*(\partial M_{\text{reg}})$ : For if  $\Phi^{\bar{t}}(x, v) \in S^*_{\partial M_{\text{reg}}} M \setminus \mathcal{S}$  (that is if the boundary were intersected transversally at a regular point) then  $\Phi^{\bar{t}+s}(x, v) \in S^*M^\circ$  for  $0 < |s| < \varepsilon_1$ .

We then claim that for  $(x', v') \in U, U$  a small neighbourhood of  $(x, v)$ , the intersections are transversal (that is, the second condition in the definition of  $X_T$  is open) and the times of intersections of  $\Phi^t(x', v')$ , with the boundary,  $t_i(x', v')$  vary continuously with  $(x', v')$ . By induction it is enough to verify this for  $t_1$ . Let us introduce normal geodesic coordinates (or any product decomposition near the boundary):  $z \mapsto (y_1(z), y'(z)), y_1(z) = d_g(\partial M_{\text{reg}}, z), y'(z) \in \partial M_{\text{reg}}$ , near  $\pi \circ \Phi^{t_1}(x, v)$ . If  $\tilde{\Phi}^s$  denotes the geodesic flow on  $S^*\tilde{M}$  then the transversality of the intersection is equivalent to  $\partial_t y_1(\tilde{\Phi}^{t_1}(x, v)) \upharpoonright_{t=t_1} \neq 0$  which is clearly an open condition. It also implies that the times of intersection with the boundary are smooth functions of  $(x', v') \in U$  by the implicit function theorem.

To see that the first condition in the definition of  $X_T$  holds for  $(x', v')$  in a neighbourhood of  $(x, v) \in X_T$ , we note that  $(x', v') \mapsto \pi \circ \Phi^{t_i(x', v')}(x', v') \in \partial M$  is continuous and hence the image of a small neighbourhood of  $(x, v)$  will avoid the closed set  $S$ .  $\square$

We will now move to analytic considerations. Since the boundary of  $M$  is Lipschitz<sup>1</sup>, the Laplace Beltrami operator,  $\Delta_g$ , with Dirichlet boundary conditions is self-adjoint on  $L^2(M, d \text{vol}_g)$  with the usual domain  $H^2(M) \cap H_0^1(M)$ . The spectrum is discrete:

$$\Delta \phi_j = \lambda_j^2 \phi_j, \quad \phi_j \upharpoonright_{\partial M} = 0, \quad 0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \rightarrow \infty. \tag{2.6}$$

We assume that the eigenfunctions are mutually orthogonal and of unit norm. We put  $N(\lambda) = \#\{j : \lambda_j \leq \lambda\}$  and have

**Lemma 3.** *For an orthonormal basis of eigenfunctions of  $\Delta_g$ ,*

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\phi_j|^2 \xrightarrow{\lambda \rightarrow \infty} 1 \quad \text{weakly in } L^1(M, d \text{vol}_g). \tag{2.7}$$

<sup>1</sup> A much weaker assumption would suffice – for this case we refer to [10] for references.

*Proof.* Let  $\tilde{\Delta}$  be the Laplacian on  $\tilde{M}$  and let  $k(t, x, y)$  and  $\tilde{k}(t, x, y)$  be the heat kernels for the Dirichlet Laplacian on  $M$  and the Laplacian on  $\tilde{M}$  respectively. A maximum principle argument of Weyl<sup>2</sup> shows that for  $x, y \in M^\circ$ ,

$$0 \leq \tilde{k}(t, x, y) - k(t, x, y) \leq \begin{cases} (4\pi t)^{-n/2} \exp(-d_g(y, \partial M)^2/4t) & 0 < t < d_g(y, \partial M)^2/C \\ (4\pi t)^{-n/2} \exp(-C/4) & d_g(y, \partial M)^2/C \leq t \end{cases}, \tag{2.8}$$

see [5] for a nice exposition. Proceeding as in that paper we then obtain  $\int_M f(x)k(t, x, x)d \text{vol}_g(x) = (4\pi t)^{-n/2} \int_M f(x)d \text{vol}_g(x) + \mathcal{O}(t^{-n/2+1/2})$ , for any  $f \in \mathcal{C}^\infty(M)$ . Representing  $k(t, x, x) = \int_0^\infty e^{-t\lambda} dN(\lambda, x)$  with  $N(\lambda, x) = \sum_{\lambda_j \leq \lambda} |\phi_j(x)|^2$ , and applying Karamata’s Tauberian theorem complete the proof.  $\square$

We will now generalize Lemma 3 to pseudodifferential operators with kernels supported away from the boundary. That assumption could be eliminated by heat equation techniques (see [6]) but we avoid that for the sake of simplicity.

**Lemma 4.** *Let  $A \in \Psi_{\text{phg}}(M)$  have the Schwartz kernel of  $A$  compactly supported in  $M^\circ \times M^\circ$ . Then with the notation of Lemma 3 and denoting by  $\sigma_0(A)$  the principal symbol of  $A$ ,*

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \langle A\phi_j, \phi_j \rangle \xrightarrow{\lambda \rightarrow \infty} \int_{S^*M^\circ} \sigma_0(A). \tag{2.9}$$

*Proof.* Since  $A$  is supported away from the boundary the lemma follows from small time asymptotics of  $\text{tr } A \exp(it\sqrt{\Delta})$  ([9], Proposition 29.1.2) which are obtained as in the boundaryless case—see [9], proof of Theorem 29.3.3 for a similar argument. Then (2.9) follows by applying Hörmander’s Tauberian Lemma (see [9], Lemma 17.5.6), precisely as in Theorem 29.1.5 of [9].  $\square$

As a final preliminary fact we need a replacement of Egorov’s Theorem for our setting. As usual we use  $WF(A)$  to denote the invariantly defined essential support of the full symbol of  $A$ .

**Lemma 5.** *Let  $A \in \Psi_{\text{phg}}(M)$  have the Schwartz kernel of  $A$  compactly supported in  $M^\circ \times M^\circ$  and in addition  $WF(A) \cap S^*M^\circ \in X_T$ . If  $\psi \in \mathcal{C}_0^\infty(M^\circ)$ , then for  $|t| < T$ ,*

$$A_t^\psi \stackrel{\text{def}}{=} \psi \exp(it\sqrt{\Delta}) \circ A \circ \exp(-it\sqrt{\Delta}) \psi \in \Psi_{\text{phg}}^0(M^\circ),$$

and

$$\sigma_0(A_t^\psi) = (\pi^*\psi)^2 \sigma_0(A) \circ \Phi^t. \tag{2.10}$$

*Proof.* By the assumption,  $|t| < T$  implies that

$$d_g(\pi \circ \Phi^t(WF(A) \cap S^*M^\circ), S) > \varepsilon_1$$

and

$$i^*(\Phi^t(WF(A) \cap S^*M^\circ)) \cap S_{\partial M_{\text{reg}}} M \cap S^* \partial M_{\text{reg}} = \emptyset,$$

<sup>2</sup> For which continuity of the boundary is a sufficient requirement.

that is the billiard flow-out of  $WF(A)$  is microlocally away from the singular and glancing sets. The lemma then follows from the “easy” (that is, non-diffractive) part of Farris’s extension of Egorov’s theorem to manifolds with diffractive boundaries [7], which in fact is already present in [2]<sup>3</sup>. We refer to those papers for the now standard details.  $\square$

### 3. Proof of the Theorem

We will prove that for all  $A \in \Psi_{\text{phg}}(M^\circ)$  with Schwartz kernels compactly supported in  $M^\circ \times M^\circ$ ,

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\langle A\phi_j, \phi_j \rangle - \overline{\sigma_0(A)}|^2 \xrightarrow{\lambda \rightarrow \infty} 0, \tag{3.1}$$

where  $\overline{\sigma_0(A)} = \int_{S^*M^\circ} \sigma_0(A) d\mu$ , where we normalize the Liouville measure so that  $\mu(S^*M^\circ) = 1$ . The existence of the subsequence satisfying (1.1) for all  $A$  will then follow by the diagonal argument used in [17, 18, 3].

To establish (3.1), we fix  $T > 0$  to be chosen arbitrarily large later, and let  $U_\varepsilon \Subset X_{2T} \setminus S^*_{\partial M_{\text{reg}}}$  be a nested family of open sets such that  $\overline{U_\varepsilon} \subset U_{\varepsilon'}$  for  $\varepsilon' < \varepsilon$  and  $\bigcup_{\varepsilon > 0} U_\varepsilon = X_{2T}$ . Then there exist operators  $E_\varepsilon \in \Psi^0_{\text{phg}}(M^\circ)$  with Schwartz kernels compactly supported in  $M^\circ \times M^\circ$  such that

$$\begin{aligned} WF(E_\varepsilon) \cap S^*M^\circ &\subset U_{2\varepsilon}, & WF(E_\varepsilon - I) \cap S^*M^\circ &\subset S^*M^\circ \setminus U_\varepsilon, \\ 0 &\leq \sigma_0(E_\varepsilon) \leq 1. \end{aligned} \tag{3.2}$$

We then define  $A_\varepsilon = E_\varepsilon A$  and study the left-hand side of (3.1) with  $A$  replaced by  $A_\varepsilon$ .

Following [18] we introduce the notation

$$\rho_j(B) = \langle B\phi_j, \phi_j \rangle, \quad \langle B \rangle_T = \frac{1}{2T} \int_{-T}^T \exp(-it\sqrt{\Delta}) \circ B \circ \exp(it\sqrt{\Delta}) dt,$$

and recall the simple observations

$$|\rho_j(B)|^2 \leq \rho_j(B^*B), \quad \rho_j(\langle B \rangle_T) = \rho_j(B). \tag{3.3}$$

Hence for  $A$  replaced by  $A_\varepsilon$ , the expression in (3.1) can be rewritten and estimated as follows:

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A_\varepsilon) - \overline{\sigma_0(A_\varepsilon)}|^2 \leq \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \rho_j(\langle A_\varepsilon - \overline{\sigma_0(A_\varepsilon)} \rangle_T^* \langle A_\varepsilon - \overline{\sigma_0(A_\varepsilon)} \rangle_T). \tag{3.4}$$

We now introduce  $\psi_i^\delta \in \mathcal{C}_0^\infty(M^\circ; [0, 1])$ ,  $i = 1, 2, 3$ , such that  $d_g(\text{supp}(\psi_i^\delta - 1), \partial M) < \delta$  and  $\psi_i^\delta = 1$  on  $\text{supp} \psi_{i+1}^\delta$ ,  $i = 1, 2$ . Writing  $B_{\varepsilon, T} = \langle A_\varepsilon - \overline{\sigma_0(A_\varepsilon)} \rangle_T^* \langle A_\varepsilon -$

<sup>3</sup> We gratefully owe this reference to Melrose.

$\overline{\sigma_0(A_\varepsilon)})_T$ , we obtain

$$\begin{aligned} \rho_j(B_{\varepsilon,T}) &= \rho_j(\psi_1^\delta B_{\varepsilon,T} \psi_2^\delta) + \rho_j((1 - \psi_1^\delta) B_{\varepsilon,T} \psi_2^\delta) \\ &\quad + \rho_j((1 - \psi_3^\delta) B_{\varepsilon,T} (1 - \psi_2^\delta)) + \rho_j(\psi_3^\delta B_{\varepsilon,T} (1 - \psi_2^\delta)). \end{aligned} \tag{3.5}$$

By the theorem on propagation of singularities along bicharacteristics intersecting the boundary transversally (that is, in the hyperbolic region – see [2] or Theorem 24.2.1 of [9]), the operators  $(1 - \psi_1^\delta) B_{\varepsilon,T} \psi_2^\delta$  and  $\psi_3^\delta B_{\varepsilon,T} (1 - \psi_2^\delta)$  have kernels in  $\mathcal{C}^\infty(M \times M)$ . Hence multiplying by  $I = (\Delta + I)^{-m} (\Delta + I)^m$  and pulling the first factor on  $\phi_j$  in the inner product defining  $\rho_j$  gives

$$\rho_j((1 - \psi_1^\delta) B_{\varepsilon,T} \psi_2^\delta), \rho_j(\psi_3^\delta B_{\varepsilon,T} (1 - \psi_2^\delta)) = \mathcal{O}_{T,\varepsilon,\delta,N}(\lambda_j^{-N}) \quad \text{for any } N.$$

The boundedness of  $B_{\varepsilon,T}$  on  $L^2(M, d \text{vol}_g)$  implies that

$$\rho_j((1 - \psi_3^\delta) B_{\varepsilon,T} (1 - \psi_2^\delta)) \leq C_{T,\varepsilon} \langle (1 - \psi_3^\delta)^2 \phi_j, \phi_j \rangle.$$

To the remaining term in (3.5) we apply Lemma 5 which combined with Lemmas 3 and 4 gives the following estimate for (3.4):

$$\begin{aligned} &\int_{S^*M} \left| \frac{1}{2T} \int_{-T}^T \pi^* \psi_2^\delta (\sigma_0(A_\varepsilon) \circ \Phi^t - \overline{\sigma_0(A_\varepsilon)}) dt \right|^2 d\mu \\ &\quad + C_{T,\varepsilon} \int_M (1 - \psi_3^\delta)^2 d \text{vol}_g + o_{\varepsilon,T,\delta}(1), \quad \text{as } \lambda \rightarrow \infty, \end{aligned} \tag{3.6}$$

and hence

$$\begin{aligned} \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A_\varepsilon) - \overline{\sigma_0(A_\varepsilon)}|^2 &\leq \int_{S^*M} \left| \frac{1}{2T} \int_{-T}^T (\sigma_0(A_\varepsilon) \circ \Phi^t - \overline{\sigma_0(A_\varepsilon)}) dt \right|^2 d\mu \\ &\quad + f_{T,\varepsilon,\delta}(\lambda) + g_{T,\varepsilon}(\delta), \end{aligned} \tag{3.7}$$

where  $f_{T,\varepsilon,\delta}(\lambda) \rightarrow 0$  as  $\lambda \rightarrow 0$  and  $g_{T,\varepsilon}(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ .

By Lemma 2,  $X_{2T}$  has full measure for all  $T > 0$  and since  $\sigma_0(E_\varepsilon) = 1$  on  $U_\varepsilon$ , we have  $\lim_{\varepsilon \rightarrow 0} \sigma_0(E_\varepsilon) = 1$  almost everywhere. Hence  $\lim_{\varepsilon \rightarrow 0} \sigma_0(A_\varepsilon) = \sigma_0(A)$  almost everywhere and hence by dominated convergence  $(\sigma_0(A_\varepsilon))$  are uniformly bounded in  $\varepsilon$

$$\begin{aligned} &\int_{S^*M} \left( \left| \frac{1}{2T} \int_{-T}^T (\sigma_0(A_\varepsilon) \circ \Phi^t - \overline{\sigma_0(A_\varepsilon)}) dt \right|^2 - \left| \frac{1}{2T} \int_{-T}^T (\sigma_0(A) \circ \Phi^t - \overline{\sigma_0(A)}) dt \right|^2 \right) d\mu \\ &= h_T(\varepsilon), \quad h_T(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0. \end{aligned}$$

On the other hand, the ergodicity of  $\Phi^t$  on a full measure subset of  $S^*M$  implies that the second term on the left-hand side above is equal to  $e(T)$ , where  $e(T) \rightarrow 0$  as  $T \rightarrow \infty$ . Going back to (3.4), we obtain

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A_\varepsilon) - \overline{\sigma_0(A_\varepsilon)}|^2 \leq e(T) + h_T(\varepsilon) + g_{T,\varepsilon}(\delta) + f_{T,\varepsilon,\delta}(\lambda), \tag{3.8}$$

which is the needed analogue of (3.1) for  $A_\varepsilon$ . Note however that  $A_\varepsilon$  depends on  $T$ .



We now claim that

$$\frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A) - \overline{\sigma_0(A)}|^2 = \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A_\varepsilon - \overline{\sigma_0(A_\varepsilon)})|^2 + \tilde{h}_{T,\varepsilon} + \tilde{f}_{T,\varepsilon}(\lambda), \tag{3.9}$$

where  $\lim_{\varepsilon \rightarrow 0} \tilde{h}_T(\varepsilon) = 0$ , and  $\lim_{\lambda \rightarrow \infty} \tilde{f}_{T,\varepsilon}(\lambda) = 0$ . In fact, putting  $R_\varepsilon = I - E_\varepsilon$  we can rewrite the left-hand side as

$$\begin{aligned} \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j((A_\varepsilon + R_\varepsilon A) - \overline{\sigma_0(A_\varepsilon + R_\varepsilon A)})|^2 &= \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} |\rho_j(A_\varepsilon - \overline{\sigma_0(A_\varepsilon)})|^2 \\ &+ \mathcal{O}(1) \left( \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \rho_j((A_\varepsilon - \overline{\sigma_0(A_\varepsilon)})^*(A_\varepsilon - \overline{\sigma_0(A_\varepsilon)})) \right)^{\frac{1}{2}} \\ &\times \left( \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \rho_j((R_\varepsilon A - \overline{\sigma_0(R_\varepsilon A)})^*(R_\varepsilon A - \overline{\sigma_0(R_\varepsilon A)})) \right)^{\frac{1}{2}} \\ &+ \mathcal{O}(1) \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \rho_j((R_\varepsilon A - \overline{\sigma_0(R_\varepsilon A)})^*(R_\varepsilon A - \overline{\sigma_0(R_\varepsilon A)})), \end{aligned}$$

where we have used the first part of (3.3) and the Schwartz inequality. Lemma 4 and the uniform boundedness of  $\sigma_0(A_\varepsilon)$  now give (3.9) since  $\sigma_0(R_\varepsilon A) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

We now combine (3.9) with (3.8) to conclude that the left-hand side of (3.1) is bounded by  $e(T) + h_T^\sharp(\varepsilon) + g_{T,\varepsilon}(\delta) + f_{T,\varepsilon,\delta}^\sharp(\lambda)$ , where  $\lim_{T \rightarrow \infty} e(T) = 0$ ,  $\lim_{\varepsilon \rightarrow 0} h_T^\sharp(\varepsilon) = 0$ ,  $\lim_{\delta \rightarrow 0} g_{T,\varepsilon}(\delta) = 0$  and  $\lim_{\lambda \rightarrow \infty} f_{T,\varepsilon,\delta}^\sharp(\lambda) = 0$ . Hence by successively choosing  $T, \varepsilon, \delta$  and  $\lambda$  we can make it arbitrarily small and that completes the proof of (3.1). To complete the proof of Theorem 1, we recall that (3.1) implies the existence of a subsequence  $\mathcal{S}_A \subset \mathbb{N}$  of density one such that

$$\lim_{j \rightarrow \infty, j \in \mathcal{S}_A} \langle A\phi_j, \phi_j \rangle = \overline{\sigma_0(A)}, \tag{3.10}$$

(see [16], Theorem 1.20). To extract a subsequence of density one independent of  $A$ , we first choose a countable set  $\{A_k\} \subset \Psi^\circ$ , with  $\|A_k\| \leq 1$  and with dense span in  $\Psi^\circ/\Psi^{-1}$ . For instance, we could take  $A_k = a_k(x, D)$  with  $\{a_k(x, \xi)\}$  a smooth orthonormal basis of  $L^2(S^*M)$ . For each  $k$ , we have then a subsequence  $\mathcal{S}_k \subset \mathbb{N}$  such that (3.10) is satisfied for  $A = A_k$ . We may assume  $\mathcal{S}_k \subset \mathcal{S}_{k+1}$  and choose  $N_k$  such that for  $N \geq N_k$ ,

$$\frac{1}{N} \#\{j \in \mathcal{S}_k : j \leq N\} \geq 1 - 2^{-k}.$$

The subsequence  $\mathcal{S}_\infty$  defined by

$$\mathcal{S}_\infty \cap [N_k, N_{k+1}] = \mathcal{S}_k \cap [N_k, N_{k+1}]$$

is then of density one and satisfies

$$\lim_{j \rightarrow \infty, j \in \mathcal{S}_\infty} \langle A\phi_j, \phi_j \rangle = \overline{\sigma_0(A)} \tag{3.11}$$

for all  $A \in \Psi^\circ$ . Indeed, (3.11) holds for all the  $A_k$ 's and hence for the closure of their linear combinations relative to the operator norm in  $\Psi^\circ$ .

*Proof of the Corollary.* This is immediate if we write  $f \in \mathcal{C}^\infty(M)$  as  $\psi f + (1 - \psi)f$  with  $\psi \in \mathcal{C}_0^\infty(M^\circ)$  – we apply the theorem to the first term and Lemma 3 to the second. We then take a sequence of  $\psi$ 's so that  $\int_M (1 - \psi) d\text{vol}_g \rightarrow 0$ .

*Acknowledgements.* We would like to thank Laurent Guillopé for his generous help during the preparation of this article and the National Science Foundation for partial support under the grants DMS-9404637 and DMS-9202344.

## References

1. Bunimovich, L.A.: On the ergodic properties of nowhere dispersive billiards. *Commun. Math. Phys.* **65**, 295–312 (1979)
2. Chazarain, J.: Construction de la paramétrie du problème mixte hyperbolique pour l'équation des ondes. *C.R. Acad. Sci. Paris* **276**, 1213–1215 (1973)
3. Colin de Verdière, Y.: Ergodicité et fonctions propres du laplacien. *Commun. Math. Phys.* **102**, 497–502 (1985)
4. Cornfeld, I.P., Fomin, S.V., Sinai, Ya.G.: *Ergodic Theory*. Grundlehren Math. Wiss. **245**, Berlin: Springer, 1982
5. Dodziuk, J.: Eigenvalues of the Laplacian and the heat equation. *Am. Math. Monthly* **88**, 686–695 (1981)
6. Gérard, P., Leichtnam, E.: Ergodic properties of eigenfunctions for the Dirichlet problem. *Duke Math. J.* **71**(2), 559–607 (1993)
7. Farris, M.: Egorov's theorem on a manifold with diffractive boundary. *Comm. P.D.E.* **6**(6), 651–687 (1981)
8. Guillemin, V., Melrose, R.B.: A cohomological invariant of discrete dynamical systems. In: Christoffel Centennial Volume, P.I. Putzer, F. Feher, eds. Basel: Birkhäuser, 1981, pp. 672–679
9. Hörmander, L.: *The Analysis of Linear Partial Differential Operators*. Vols. **3** and **4**, Grundlehren Math. Wiss. **274** and **275**, Berlin, Heidelberg, New York: Springer, 1986
10. Kenig, C., Pipher, J.: The  $h$ -path distribution of the lifetime of conditioned Brownian motion for non-smooth domains. *Probability Th. Rel. Fields* **82**, 615–623 (1989)
11. Kerckhoff, S., Masur, H., Smillie, J.: Ergodicity of billiard flows and quadratic differentials. *Ann. Math.* **124**, 293–311 (1986)
12. Melrose, R.B., Sjöstrand, J.: Singularities in boundary value problems I. *Comm. Pure Appl. Math.* **31**, 593–617 (1978)
13. Petkov, V., Stoyanov, L.N.: *Geometry of Reflecting Rays and Inverse Spectral Problems*. New York: John Wiley and Sons, 1992
14. Schnirelman, A.I.: Ergodic properties of eigenfunctions. *Usp. Math. Nauk* **29**, 181–182 (1974)
15. Sinai, Ya.G.: Dynamical systems with elastic reflections. Ergodic properties of dispersive billiards. *Usp. Mat. Nauk* **25**, 141–192 (1970), *Russ. Math. Surv.* **25**, 137–189 (1970)
16. Walters, P.: *An Introduction to Ergodic Theory*. Grad. Texts in Math. **79**, Berlin, Heidelberg, New York: Springer, 1982
17. Zelditch, S.: Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.* **55**, 919–941 (1987)
18. Zelditch, S.: Quantum ergodicity of  $C^*$  dynamical systems, to appear in: *Commun. Math. Phys.*